

**Sections of a Stage 1 Registered Report**

**Scientific Repo**

[Abstract 1](#_Toc126850340)

[Introduction 2](#_Toc126850341)

[Methods 3](#_Toc126850342)

[Ethics information 3](#_Toc126850343)

[Secondary Data Description 3](#_Toc126850344)

[Data Collection Procedures 4](#_Toc126850345)

[Experimental design 4](#_Toc126850346)

[Measured Variables 4](#_Toc126850347)

[Stimuli and materials 4](#_Toc126850348)

[Procedure 5](#_Toc126850349)

[Pilot data simulated data 5](#_Toc126850350)

[Analysis Plan 6](#_Toc126850351)

[Data pre-processing 6](#_Toc126850352)

[Calculation of indices & quantifying SPE in the SALT 6](#_Toc126850353)

[Reliability of indices in SALT as individual-level/group-level 7](#_Toc126850354)

[Effect related to practice in SALT 8](#_Toc126850355)

[Split-half reliability of SPE in SALT 8](#_Toc126850356)

[Data availability 9](#_Toc126850357)

[Code availability 9](#_Toc126850358)

[Results 9](#_Toc126850359)

[Descriptive Statistics 9](#_Toc126850360)

[ICC(Intraclass correlation coefficient) 10](#_Toc126850361)

[HLM(Hierarchical Linear Model) 10](#_Toc126850362)

[Split-Half Reliability 11](#_Toc126850363)

[Discussion 12](#_Toc126850364)

[Acknowledgements 13](#_Toc126850365)

[Author contributions 13](#_Toc126850366)

[Competing interests 13](#_Toc126850367)

[Figures 13](#_Toc126850368)

[Figure Legends 13](#_Toc126850369)

[Supplementary information 14](#_Toc126850370)

[References 17](#_Toc126850371)

**Estimating Reliability of the Self-Associative Learning Task as a Measure of Self-Prioritization Effect: Re-analyses of a Longitudinal Dataset**

Zheng Liu 1, Yuanrui Zheng2, Jie Sui3\*, Hu Chuan-Peng 1\*

1 School of Psychology, Nanjing Normal University, Nanjing, China

2 School of Education, Kunming City College, Kunming, China

3 School of Psychology, University of Aberdeen, Old Aberdeen, Scotland

\* Corresponding authors: Hu Chuan-Peng ([hu.chuan-peng@nnu.edu.cn](mailto:hu.chuan-peng@nnu.edu.cn))

# **Abstract**

In recent years, the Self-Associative Learning Task (SALT) has been a valuable tool in studying the regulation of self-related information in individual cognitive processing. However, there has been limited research on the psychometric properties of SALT's outcomes. Additionally, while SALT is a relatively simple task, there are multiple ways to measure the self-prioritization effect, such as through reaction time-based and accuracy-based indices. It is unknown whether these indices accurately capture the self-prioritization effect, and if so, which index is the most reliable at both the group and individual level. To address these questions, we plan to reanalyze a longitudinal dataset collected in 2016 that involved 34 healthy volunteers who completed the SALT task over six sessions, separated by one week. By using intraclass correlations, multilevel modeling, and split-half reliability, we aim to conduct a comprehensive examination of the test-retest reliability and potential practice effect in SALT. This study will provide important insights into SALT and pave the way for its use in further research, clinical applications, and personal performance monitoring.

# **Introduction**

Self-bias has been found to be a consistent phenomenon across various cognitive domains, including perception, attention, memory, and decision-making (Cunningham & Turk, 2017; Desebrock et al., 2018; Sui & Humphreys, 2013)The **Self-Prioritization Effect (SPE)** has long been established as a phenomenon where people remember information better when it is related to themselves compared to information related to others(Rogers et al., 1977; Symons & Johnson, 1997). This is because the self-concept is so fundamental that any arbitrary stimulus linked to it can quickly become significant and impact perception (Humphreys & Sui, 2015; Sui et al., 2015).

To measure the SPE, various tests have been used, such as the trait-adjectives paradigm(Craik & Tulving, 1975; Rogers et al., 1977), attentional blink paradigm(Shapiro et al., 1997), and the ownership task (Cunningham et al., 2008)(see a review on(Amodeo et al., 2021)). However, it is difficult to determine if self-related stimuli, like the participant's own name or face, are processed faster because they are associated with the self or because they are more familiar. **The self-associative learning task (SALT)** developed by Sui and colleagues provides an effective approach to studying self-related information in individual cognitive processing, avoiding the confound of familiarity(Sui et al., 2012). In the SALT, participants first associate geometrical shapes (e.g., triangle, square, and circle) with labels (e.g., "You," "friend," and "stranger"). In the second stage, participants decide if the shape-label pairings match. Typically, a significant SPE is found with faster response times, better accuracy, and higher sensitivity scores for self-shapes compared to friend and stranger shapes(Schäfer & Frings, 2019; Sel et al., 2019; Sui et al., 2016). The drift-diffusion model was also used to demonstrate the perceptual matching of immediately acquired self-relatedness, showing faster evidence accumulation (drift rate) and a bias at the start of information accumulation (starting point, z)(Golubickis et al., 2017; Macrae et al., 2017; Yankouskaya et al., 2020). These effects have been robustly replicated in multiple studies.

The use of the self-associative learning paradigm has increased greatly in recent years, due to its convenience in studying powerful top-down processing and avoiding the confounding influence of stimuli familiarity. Psychologists in the fields of clinical health and mental illness have utilized the paradigm to understand atypical self-processing in populations such as those with autism or depression (Gillespie‐Smith et al., 2018; Nijhof & Bird, 2019; Sui & Humphreys, 2017). It has also been used to examine group processes and cultural differences(Jiang et al., 2019). and even adapted for use with children to study the development of self-advantage(Maire et al., 2020).

Despite the widespread use of the SALT, there has been a lack of detailed examination and reporting of the psychometric properties of its outcomes. This is crucial, especially if the self-associative learning paradigm is to be used in clinical settings, such as diagnosing depression(Liu et al., 2022). For accurate assessment of human perceptual abilities, cognitive tests must have high reliability, meaning consistency in their measurements(Parsons et al., 2019). However, there are multiple ways to quantify the self-prioritization effect in a task as simple as the SALT, and it is currently unclear (1) whether these indices consistently capture the self-prioritization effect over time, and if so, (2) which index is most suitable for repeated measurements.

Our research aims to examine the reliability and stability of commonly used indices for measuring self-prioritization effect (SPE) in the Self-Associative Learning Task (SALT). To achieve this, we will re-analyze a pre-existing dataset, where participants associated three shapes with labels for themselves, a friend, or a stranger, over six testing sessions with one-week intervals.

Our research has three main goals:

1. *Test which index (s) is appropriate and consistent to indicate the group-level self-prioritization effect (SPE) in the SALT;*
2. *Test which index (s) is consistent to indicate the individual-level self-prioritization effect (SPE) in the SALT;*
3. *Test whether there is a practice effect across testing sessions.*

Our main hypothesis are as follows:

1. *(a) Model-based measurements and Reaction time-based measurement are appropriately reliable as group-level SPE indicators in the associative learning task (b) accuracy-based measurement exhibits different degrees of inconsistency from one time point to another.*
2. *(a) Model-based measurements, which reflect the critical underlying generative process of individuals, are appropriately reliable as individual-level SPE indicators in the associative learning task. (b) RT and accuracy-based measurements exhibits different degrees of inconsistency from one time point to another.*
3. *(a) There is a practice effect on all indices across testing sessions.*

We aim to test our hypotheses using Hierarchical Linear Model (HLM), Intraclass Correlation Coefficient (ICC), and Split-Half Reliability. The results of this study will provide valuable insights into the reliability and consistency of the Self-Associative Learning Task (SALT), which could pave the way for its future use in research, clinical settings, and personal performance monitoring. For more information, see our Analysis Plan.

# **Methods**

## Ethics information

Our research does not involve any treatment of humans or animals and is a secondary analysis of pre-existing data. As such, informed consent and confidentiality are not relevant. The original study from which the data was collected was approved ethically by the research committee at Tsinghua University.

## Secondary Data Description

To address our research questions, we'll use a pre-existing dataset from a study conducted by Hu Chuan-Peng at Tsinghua University in 2016. The original study aimed to compare the self-prioritization effect (SPE) between sub-clinical depressed and non-depressed participants, but only the non-depressed group was collected due to difficulty in recruiting sub-clinical depressed participants. The dataset contains data from 34 non-depressed and 6 depressed participants, who participated in six testing sessions over a 1-week interval. Each session included a modified SALT task, a set of questionnaires, and another modified SALT task. We plan to **use the results of the neutral condition in the second SALT task from the** **34[[1]](#footnote-1)**  **non-depressed participants with relatively low scores on the depression-related questionnaire.**

## Data Collection Procedures

36 college students from the Tsinghua University community participated in the experiment and received compensation. All participants were right-handed and had normal or corrected-to-normal vision. Unfortunately, data from one participant was excluded due to confusing participant information provided to the experimenter, and data from one male participant was missing due to a programming error. This left a total of 34 valid participants, with 21 females and 13 males, averaging 21 years old (SD = 2.52) in age.

## Experimental design

Experiment B is a four-factor design, with 2 levels of match vs. non-match, 3 levels of identity (self, friend, stranger), 4 levels of emotion (control, neutral, happy, sad), and 6 repeated sessions. Its purpose is to examine the self-bias effect under different emotions (happy, sad, neutral, control).

## Measured Variables

At each wave, participants' keypress, reaction time, and accuracy in each trial were recorded. The participants also filled out questionnaires that varied from wave to wave and covered topics such as personal wellbeing, physical and mental health, and psychological distance between the self, a friend, and a stranger.

## Stimuli and materials

The experiment was conducted individually in a dimly lit room, using E-Prime 2.0 software on a PC with a 1024 x 768 resolution monitor, refreshing at 100 Hz. Participants recorded their keypresses, reaction time, and accuracy during each trial.

The experiment was split into two phases. The first phase followed the study by (Sui et al., 2012)and involved a learning task where participants paired geometric shapes with labels. The shapes were not shown at this stage. The learning task lasted approximately 60 seconds, and the shape-label associations were balanced across participants. Then, in the matching task, a fixation cross was displayed in the center of the screen for 500 ms, followed by the presentation of a shape-label pairing and the fixation cross for 100 ms. Then, the screen went blank for 1500 ms, or until a response was made. Participants were asked to determine whether the shape matched the label by pressing one of two buttons as quickly and accurately as possible within this timeframe.

The participants took part in a two-phase experiment. In the first phase, they learned four sets of associations between shapes and labels, with one set being a control condition and three others being emotion-based. The control condition involved associating three geometric shapes (circle, horizontal ellipse, and vertical ellipse) with three labels (self, friend, and stranger), while the emotion-based conditions showed facial expressions (happy, sad, and neutral) on the shapes. Before starting the formal trials, each participant went through a training session with 24 practice trials. After the training, each participant completed 6 blocks of 60 trials in the matching task, with 2 match types (match/mismatch) × 3 shape associations, for a total of 60 trials per association. Participants had a short break after each block, lasting up to 60 seconds.

Diagram

Description automatically generated

**Figure 2.** In Experiment B, the stimuli and procedure were carried out in Chinese. Participants learned to associate four sets of shapes with labels, including one control condition and three emotion-based conditions. During the learning task, the shape-label matches were evenly distributed among participants, and no feedback was given during the formal trials. The example illustrates the timeline of the experiment.

## Procedure

Participants were given informed consent and took part in 80-minute experiments that included behavioral experiment A and B, as well as questionnaires. They repeated the same experiment five times in the following five weeks. They filled out questionnaires like the Beck Depression Inventory (BDI)(王振 et al., 2011), Beck Anxiety Inventory (BAI), Positive and Negative Affect Scale (PANAS)(王力 et al., 2007), and State Self-Esteem Scale(Heatherton & Polivy, 1991), and the psychological distance between self, friend, and stranger. Additionally, participants took the big-five, Rosenberg Trait self-esteem, IPA (locus of control) (Levenson, 1974; 汪向东 et al., 1999)and Belief in Free Will (Paulhus & Carey, 2010) questionnaires at the first and last sessions. All materials were presented in Chinese.

## Pilot data simulated data

To avoid any potential biases in hypothesis formation, we didn't conduct any statistical analysis on the primary data during stage 1 registration. Instead, we generated a pilot dataset with the same format as the primary data. We used an open dataset from a previous study examining the self-prioritization effect as a reference to create our pilot data.

We utilized Bootstrap methods, drawing samples from Hu et al. (2020) open dataset (accessible at <https://osf.io/mhdsn/>) with replacement (allowing the same sample to be repeated in the pilot data). The pilot data includes 6 sessions of data from 34 participants, with each participant having 24 practice trials and 360 experimental trials (6 different types of shape-label associations: two matches (matched/mismatched) x three identity associations (self, friend, stranger), 60 trials per association) per session. Figure 1 shows the first 6 rows of the pilot data.

Table

Description automatically generated

Figure 1. The first six rows of the pilot data

We ran the pilot data through our proposed statistical analysis to see whether our proposed analysis is appropriate for the secondary data structure (see analysis plan).

## Analysis Plan

To analyze our behavioral data, we'll use HDDM, a Python toolkit for Bayesian Hierarchical Modeling(Wiecki et al., 2013), and fit the data into the drift diffusion model (DDM). We'll also use the R Project(Team, 2010).

### Data pre-processing

First, we will pre-process the secondary data using the following criteria (we do not pre-process the secondary data at stage 1 registration):

1. Participant exclusion criteria
2. Participant who has the wrong trial numbers because of procedure errors should be excluded from the analysis.
3. Participants with an overall accuracy < 0.5 should be excluded from the analysis.
4. Participants with any of the conditions with zero accuracy should be excluded from the analysis.
5. Behavioural data exclusion criteria
6. Trials with no response or wrong key press should be excluded from the analysis.
7. Trials with responses less than 200 ms or faster than 1500 ms should be excluded from the analysis.
8. The practice trials will be excluded from the formal analysis.
9. The data under conditions other than the “control condition” will not be used in the current study.

### Calculation of indices & quantifying SPE in the SALT

Next, we'll calculate various metrics in the SALT and assess the Self-Prioritization Effect (SPE) at the individual level. We'll use seven common metrics for this purpose. Table 2 outlines how these metrics are calculated, as well as how the SPE is determined from them.

Table 2. Indices in SALT and corresponding calculation of indices and SPE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Indices ID** | **Indices Calculation** | **SPE Calculation Based on Indices** | | **Source** |
| Mean Reaction times (RT) |  | Type 1 calculation | Self-match - other-match | Sui et al. (2012) |
| Type 2 calculation | self-all - other-all | Sui et al. (2012) |
| Accuracy (ACC) |  | self-match) - other-match | | Sui et al. (2012) |
| d-prime | z-score (ACC (match) - z-score (1 - ACC (non-match)) | self - other | | Sui et al. (2012) |
| Efficiency |  | self-match - other-match | | Humphreys and Sui (2015); (Stoeber & Eysenck, 2008) |
| Drift rate (v) | DDM：parameters will be identified through model selection | self-match- other-match | | Golubickis et al. (2017) |
| Starting point (z) | self-match - other-match | | Golubickis et al. (2017) |

*Note.* DDM =drift diffusion model.

We'll present the average and standard deviation for each index for each session, along with other important descriptive statistics.

### Reliability of indices in SALT as individual-level/group-level

We'll assess the reliability of the SALT indices using the Intraclass Correlation Coefficient (ICC). ICC is a well-established measure of reliability in test-retest, intra-rater, and inter-rater studies(Fisher, 1992). Compared to Pearson correlation coefficient, ICC considers both the degree of correlation and agreement between multiple measurements, making it a more comprehensive measure of test-retest reliability(Koo & Li, 2016).

Specifically, we will use two-way single-measurement mixed model with absolute agreement between scores of six session (ICC2k) as the reliability measure of group-level SPE across six sessions. For the calculation of ICC2k estimates and their 95% confidence intervals, the formula is:

*Note.* = mean square for rows; = mean square for error; = mean square for columns; = number of subjects; = number of raters/measurements.

We will use a two-way multiple rater’s random effect model with absolute agreement between scores of six sessions (ICC2 ) as the reliability measure of individual-level SPE across six sessions. For the calculation of ICC2 estimates and their 95% confidence intervals, the formula is:

*Note.* = mean square for rows; = mean square for error; = mean square for columns; = number of subjects.

We'll interpret the ICC2 and ICC2k following these guidelines: a value less than 0.6 means poor reliability, a value between 0.6 and 0.8 indicates substantial reliability, and a value greater than 0.8 means excellent reliability(Cicchetti & Sparrow, 1981; Kupper & Hafner, 1989).

### Effect related to practice in SALT

To explore the potential effect of practice, we'll use a Hierarchical Linear Model with restricted maximum likelihood estimates. We'll consider sessions as fixed effects and include a random intercept to account for inter-individual differences in baseline performance. This multilevel modeling approach allows us to compare multiple sessions and consider additional predictors, such as the number of previous sessions(Ding & Vancleef, 2022). For instance, we may see a decrease in disparity between results with more practice on the test.

We will construct the hierarchical model for each index. The hierarchical model specification was as follows:

Significance will be calculated using the lmerTest package in R(Kuznetsova et al., 2017), which applies Satterthwaite’s method to estimate degrees of freedom and generate *p*-values for mixed models.

A detailed description of correspondence between each hypothesis, each statistical test and interpretation of results is illustrated in the Design Table.

### Split-half reliability of SPE in SALT

In psychological research, Cronbach's alpha is often used to determine the reliability of experiments. However, using this method in cognitive experiments can lead to biased results. As a result, more and more studies are using split-half reliability instead of Cronbach's alpha to assess the reliability of cognitive experiments. This is because Cronbach's alpha is calculated based on different experimental conditions, while split-half reliability is calculated based on trial sequences. (Kahveci et al., 2022)

There are four types of split-half reliability: odd-even, front-back, permutation, and Monte Carlo. The odd-even split separates trials into odd and even numbered sequences, while the front-back split separates the first and second halves of trials. The permutation split shuffles the trial order and randomly assigns each half to a group. Monte Carlo split-half is similar to the permutation split-half, repeating the process thousands of times to calculate the average and 95% confidence interval of the split-half reliability. This study will primarily use Monte Carlo split-half to determine the split-half reliability of SALT.

First, the data will be stratified according to Session, Match, and Identity. If not stratified, directly splitting the data in half will result in uneven distribution of trials for each experimental condition in the two halves, thereby overestimating or underestimating the reliability of the split. Therefore, after the data is stratified, we split the data. For example, when using Monte Carlo Split-Half, we randomly split the data into two half. Then we repeat this operation 1000 times. This will result in 1000 pairs of two halves of the data. Next, we use these 1000 pairs of data to calculate 1000 Pearson correlation coefficients, and then obtain the average and 95% confidence interval of the Monte Carlo split reliability. As for first-second split, odd-even split, and permutated split, they are similar to Monte Carlo division, but they only perform one split, so only one split-half reliability is obtained without interval estimate of the split-half reliability.

# **Data availability**

We will adhere to the following open science practices: open materials, open data. We will share the raw data, excluding sensitive participants’ information on acceptance of our Stage 2 manuscript. The simulated data is accessible on the Open Science Framework () and GitHub ().

# **Code availability**

Code used to simulate and analyze the pilot data is made accessible in the same location: Open Science Framework () and GitHub ().

# **Results**

## Descriptive Statistics

As shown in Table 1, we performed descriptive statistics on the six indicators for each Sessions.

Table 3 Descriptive Statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Session 6 |
| RT(ms) | 3.96(3.11) | 7.1(31.61) | 3.42(26.81) | -1.67(26.38) | -2.74(21.61) | 4.67(21.78) |
| ACC | 0(.05) | -.01(.06) | -.01(.06) | -.01(.05) | .01(.08) | 0(.06) |
| D-prime | .02(.33) | -.01(.42) | -.04(.25) | -.04(.38) | .06(.39) | .02(.32) |
| Efficiency | 2.79(58.5) | 18.14(75.16) | 1.19(63.51) | 9.66(62.48) | -7.03(85.87) | 9.46(69.07) |
| v(ms) | -57.82(2.65) | -74.95(2.8) | 52.16(2.91) | 37.22(2.55) | -47.73(2.05) | -.19(2.21) |
| z(ms) | 1.12(.67) | 3.63(1.13) | -9.98(.89) | -2.96(.88) | 4.59(.77) | -3.7(.73) |

RT reaction time, ACC accuracy, v drift rate, z starting point

## ICC(Intraclass correlation coefficient)

Intraclass correlation coefficient (ICC) is a measure of the consistency or reliability of measurements made by different raters (observers) or repeated measurements made by the same rater (observer). In essence, it tells us how much of the variation in the data can be attributed to differences between raters or repeated measurements, and how much of it can be attributed to differences within the subjects being measured. In simple terms, it gives an idea of the proportion of total variation in the data that is due to the true differences between subjects, versus due to measurement error or random fluctuations.

The present study aimed to investigate the stability of six indices, including reaction time (RT), accuracy (ACC), Dprime, Efficiency, drift rate (v) and starting point (z) in the diffusion decision model (DDM), across six time sessions. We use the Intraclass Correlation Coefficients (ICC) to evaluate how much of the variation in SALT can be attributed to within-subject repeatability over time, and how much can be attributed to between-subject differences. Among them, we are most interested in ICC2 and ICC2k, where ICC2 represents the ratio of between-subject variance to total variance, and ICC2k represents the ratio of within-subject variance to total variance. Therefore, we want ICC2 to be as large as possible and ICC2k to be as small as possible, indicating that the differences in our experimental measures are mainly due to between-subject individual differences, and each subject's performance is relatively stable across the six sessions. As shown in Figure 1, the ICC2 values of the six indices are relatively large and ICC2k values are relatively small, supporting our hypothesis.

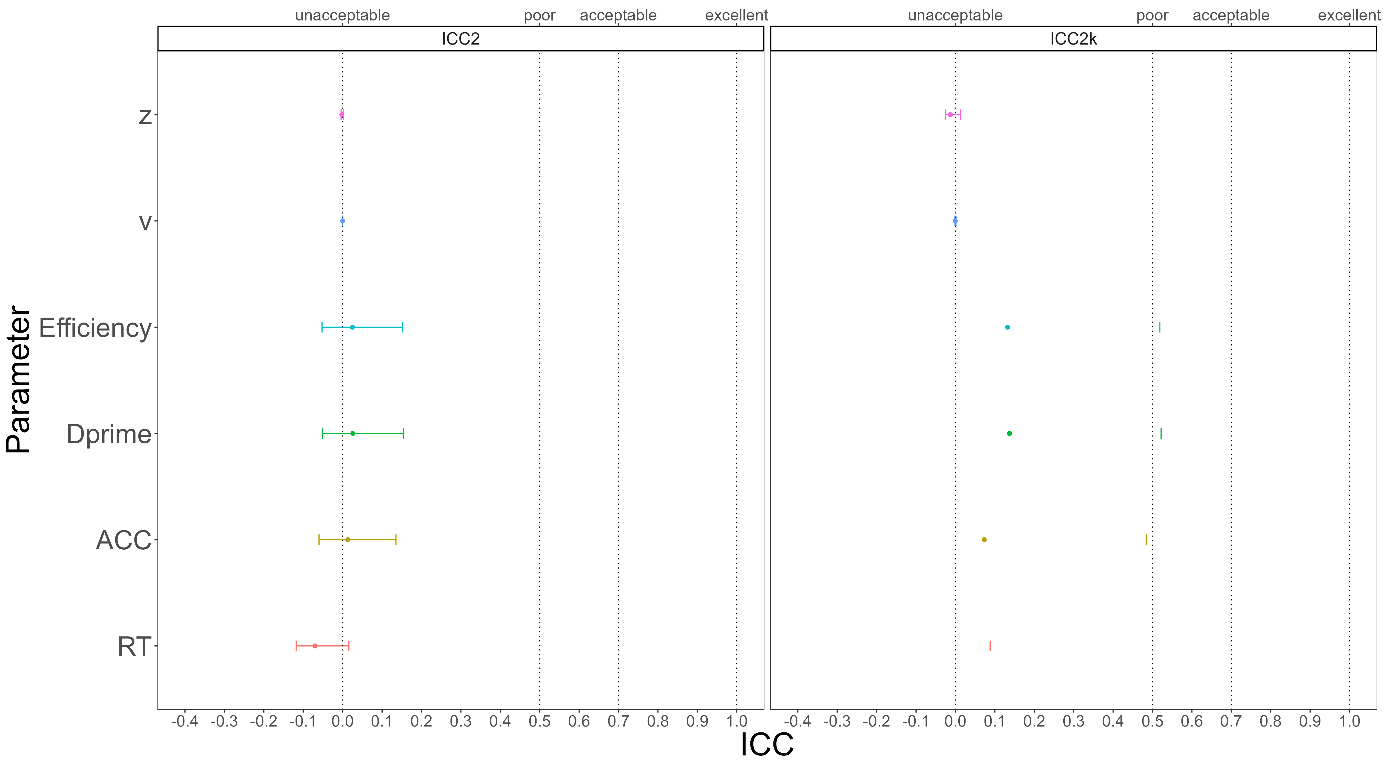


Figure. 2 ICC2 and ICC2k

## HLM(Hierarchical Linear Model)

The HLM equation for the reaction time (RT\_ms) in the study is as follows: RT\_ms ~ Session\*Match\*Identity + (1|Subject). In this equation, Match and Identity are two independent variables, and Session represents time. In our hypothesis, the results of the SALT experiment should be temporally stable. If the results of the HLM meet our hypothesis, then the variance explained by the Session should be as small as possible, while the variance explained by the Match and Identity should be as large as possible. In the between-subject variance, it should be mainly explained by the Subject. In other words, the conclusion that HLM hopes to get is similar to ICC, and we hope to prove through these two methods that the results of the SALT experiment are temporally stable and that the differences in reaction time are mainly due to individual differences among the subjects.

In our results, the variance between subjects is primarily explained by the subjects themselves, with a regression coefficient of XXX, explaining XXX% of the between-subject variability. The within-subject variance is primarily explained by the experiment variables Match and Identity, with regression coefficients of XXX and XXX respectively, while the regression coefficient of Session is small, XXX. In the HLM results, the reaction time variability is divided into within-subject variability and between-subject variability, with the within-subject variability primarily explained by the two experiment variables Match and Identity. The between-subject variability is primarily explained by the differences between subjects. The variance explained by the subjects themselves is XXX, far greater than the variance explained by Session, XXX. Therefore, the HLM results support our hypothesis that the results of the SALT experiment are stable across time, and that differences in reaction time are primarily due to individual differences and the experiment variables Match and Identity.

## Split-Half Reliability

First, we stratified the data based on three variables: Session, Match, and Identity, and then split the stratified data into two halves using four methods. Next, we calculated the SPE for each of the six indices for each half of the data. Finally, we calculated the split-half reliability for each of the six SPEs. As shown in Figure 2, when using the Monte Carlo split-half, the split-half reliability of the six indices obtained is very high, with the highest value of XXX, which means that it is the most stable of the six SPE indexing calculations for half-confidence. The results from the other three split-half methods were similar to the Monte Carlo method, which will be presented in the supplementary material.

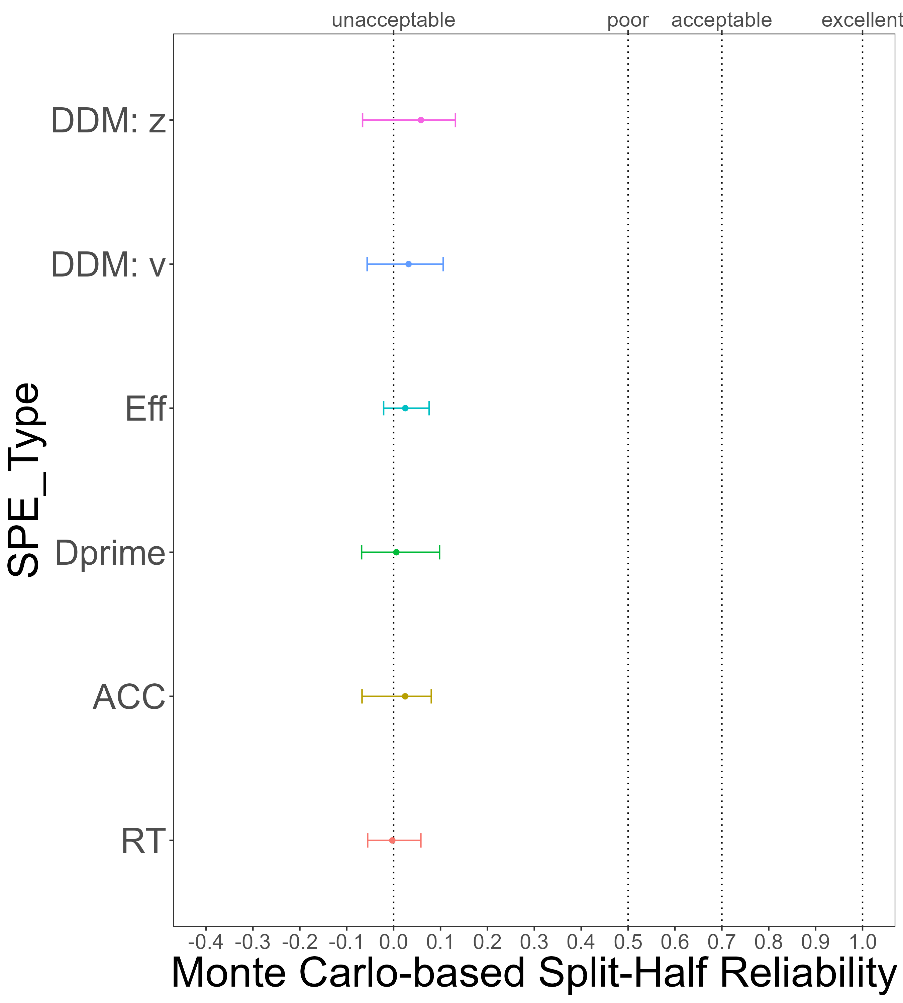


Figure. 2 Monte Carlo-based Split-Half Reliability

# **Discussion**

Do **not** include a **Discussion** section.

# **Acknowledgements**

The present research is support by xxx.

# **Author contributions**

HCP contributed to the conception and supervision of the study and will provide the methodology expertise. JS contributed to fund raising, HCP contributed to data collection. ZL and ZYR will perform the data pre-processing, analysis and visualize the results. In addition, ZL, JS and HCP will contribute to discussing the results and the drafting of the final manuscript. All authors will critically revise the manuscript.

# **Competing interests**

The authors declare no competing interests.

# **Figures**

You are encouraged to include Figures in the text or at the end of the protocol. Keep in mind that a total of 8 display elements (i.e., combination of Tables and Figures) is permitted in the final, Stage 2, submission. However, to enable typesetting of papers, we advise making the number of display items commensurate with your overall word length (that is, for a shorter paper the number of display items should be lower, for a longer manuscript a higher number may be allowed). Figures/Tables that are not essential should be included in your Supplementary Information file.

# **Figure Legends**

**Figure 1. Guidelines for the preparation of figure captions.** Figure captions should be concise. Begin with a brief title and then describe what is presented in the figure and detail all relevant statistical information. If you show pilot data, list the N of each plot and report full statistics. Aim not to exceed 350 words per legend.

# **Supplementary information**

Please report pilot data in detail here and include any other material that provides background information.

Supplementary Table 1 Design Table

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Hypothesis (if applicable)** | **Analysis Plan** | **Interpretation given to different outcomes** |
| Which indicator (s) is appropriate and consistent to indicate the group-level self-prioritization effect (SPE) in the SALT? | (a) Model-based measurement (*v, z*) and Reaction time-based measurements (Mean Reaction times) are appropriately reliable as group-level SPE indicators in the associative learning task (b) accuracy-based measurements (accuracy, d-prime, efficiency) exhibits different degrees of inconsistency from one time point to another. | We will use two-way single-measurement mixed model with absolute agreement between scores of six session (ICC2k) as reliability measure of group-level SPE across six sessions. | ICC 2k values less than 0.6 are indicative of poor reliability, values between 0.6 and 0.8 indicate substantial reliability, values greater than 0.8 indicate excellent reliability. |
| Which indicator (s) is appropriate and consistent to indicate the individual-level self-prioritization effect (SPE) in the SALT? | (a) Model-based measurement (*v, z*), which may reflect the critical underlying generative process of individuals, are appropriately reliable as individual-level SPE indicators in the associative learning task. (b) RT and accuracy-based measurements (Mean Reaction times, accuracy, d-prime, efficiency) exhibit different degrees of inconsistency from one time point to another. | We will use a two-way multiple raters random effect model with absolute agreement between scores of six sessions (ICC2) as reliability measure of individual-level SPE across six sessions. | ICC 2 values less than 0.6 are indicative of poor reliability, values between 0.6 and 0.8 indicate substantial reliability, values greater than 0.8 indicate excellent reliability. |
| Is there a practice effect across testing sessions? | There is a practice effect on all indices across testing sessions. | The effect of practice will be explored using hierarchical modelling using restricted maximum likelihood estimates with sessions as fixed effects and a random intercept to account for inter-individual differences in baseline performance. Significance will be calculated using Satterthwaite’s method to estimate degrees of freedom and generate *p-*values for mixed models. | *p*<0.05 as evidence for the presence of a practice effect. |

Supplementary Table 2 Split-Half Reliability of Other Split Method

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SPE\_type | SH\_type | SH\_r |  | SPE\_type | SH\_type | SH\_r |
| RT | First-Second | .01 |  | Efficiency | First-Second | .07 |
| RT | Odd-Even | -.05 |  | Efficiency | Odd-Even | -.04 |
| RT | Permuted | .01 |  | Efficiency | Permuted | .05 |
| ACC | First-Second | .02 |  | DDM: v | First-Second | .04 |
| ACC | Odd-Even | -.05 |  | DDM: v | Odd-Even | -.05 |
| ACC | Permuted | .07 |  | DDM: v | Permuted | .10 |
| Dprime | First-Second | .01 |  | DDM: z | First-Second | .07 |
| Dprime | Odd-Even | -.08 |  | DDM: z | Odd-Even | .01 |
| Dprime | Permuted | -.02 |  | DDM: z | Permuted | .13 |

# **References**

Amodeo, L., et al. (2021). A comparison of self-bias measures across cognitive domains. *BMC Psychology, 9*(1), 1-132. <https://doi.org/10.1186/s40359-021-00639-x>

Cicchetti, D.V., & Sparrow, S.A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic, 86*(2), 127-137.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294. <https://doi.org/10.1037/0096-3445.104.3.268>

Cunningham, S.J., & Turk, D.J. (2017, Jun). Editorial: A review of self-processing biases in cognition. *Quarterly journal of experimental psychology, 70*(6), 987-995. <https://doi.org/10.1080/17470218.2016.1276609>

Cunningham, S.J., et al. (2008). Yours or mine? Ownership and memory. *Consciousness and cognition, 17*(1), 312-318. <https://doi.org/10.1016/j.concog.2007.04.003>

Desebrock, C., et al. (2018, Oct). Self-reference in action: Arm-movement responses are enhanced in perceptual matching. *Acta psychologica, 190*, 258-266. <https://doi.org/10.1016/j.actpsy.2018.08.009>

Ding, X., & Vancleef, K. (2022). Test–retest reliability and practice effect of the Leuven Perceptual Organisation Screening Test. *Behavior Research Methods, 54*(5), 2457-2462. <https://doi.org/10.3758/s13428-021-01741-z>

Fisher, R.A. (1992). Statistical methods for research workers. *Springer New York*.

Gillespie‐Smith, K., et al. (2018). The I in autism: Severity and social functioning in autism are related to self‐processing. *British journal of developmental psychology, 36*(1), 127-141. <https://doi.org/10.1111/bjdp.12219>

Golubickis, M., et al. (2017). Self-prioritization and perceptual matching: The effects of temporal construal. *Mem Cognit, 45*(7), 1223-1239. <https://doi.org/10.3758/s13421-017-0722-3>

Heatherton, T.F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*(6), 895-910. <https://doi.org/10.1037/0022-3514.60.6.895>

Hu, C.-P., et al. (2020). Good Me Bad Me: Prioritization of the Good-Self During Perceptual Decision-Making. *Collabra. Psychology, 6*(1), 20. <https://doi.org/10.1525/collabra.301>

Humphreys, G.W., & Sui, J. (2015). The salient self: Social saliency effects based on self-bias. *Journal of cognitive psychology, 27*(2), 129-140. <https://doi.org/10.1080/20445911.2014.996156>

Jiang, M., et al. (2019). Cultural Orientation of Self-Bias in Perceptual Matching. *Front Psychol, 10*, 1469. <https://doi.org/10.3389/fpsyg.2019.01469>

Kahveci, S., et al. (2022). Reliability of reaction time tasks: how should it be computed? *Preprint*, 1-30. <https://doi.org/10.31234/osf.io/ta59r>

Koo, T.K., & Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine, 15*(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>

Kupper, L.L., & Hafner, K.b. (1989). On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics, 45*(3), 957-967. <https://doi.org/10.2307/2531695>

Kuznetsova, A., et al. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of statistical software, 82*(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>

Levenson, H. (1974). Activism and Powerful Others: Distinctions within the Concept of Internal-External Control. *Journal of Personality Assessment, 38*(4), 377-383. <https://doi.org/10.1080/00223891.1974.10119988>

Liu, Y.S., et al. (2022). Depression screening using a non-verbal self-association task: A machine-learning based pilot study. *Journal of Affective Disorders, 310*, 87-95. <https://doi.org/10.1016/j.jad.2022.04.122>

Macrae, C.N., et al. (2017). Self-Relevance Prioritizes Access to Visual Awareness. *Journal of experimental psychology. Human perception and performance, 43*(3), 438-443. <https://doi.org/10.1037/xhp0000361>

Maire, H., et al. (2020). A Developmental Study of the Self‐Prioritization Effect in Children Between 6 and 10 Years of Age. *Child development, 91*(3), 694-704. <https://doi.org/10.1111/cdev.13352>

Nijhof, A.D., & Bird, G. (2019). Self‐processing in individuals with autism spectrum disorder. *Autism research, 12*(11), 1580-1584. <https://doi.org/10.1002/aur.2200>

Parsons, S., et al. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in methods and practices in psychological science, 2*(4), 378-395. <https://doi.org/10.1177/2515245919879695>

Paulhus, D.L., & Carey, J.M. (2010). The FAD–Plus: Measuring Lay Beliefs Regarding Free Will and Related Constructs. *Journal of Personality Assessment, 93*(1), 96-104. <https://doi.org/10.1080/00223891.2010.528483>

Rogers, T.B., et al. (1977, Sep). Self-reference and the encoding of personal information. *J Pers Soc Psychol, 35*(9), 677-688. <https://doi.org/10.1037//0022-3514.35.9.677>

Schäfer, S., & Frings, C. (2019). Understanding self-prioritisation: the prioritisation of self-relevant stimuli and its relation to the individual self-esteem. *Journal of cognitive psychology, 31*(8), 813-824. <https://doi.org/10.1080/20445911.2019.1686393>

Sel, A., et al. (2019). Self-Association and Attentional Processing Regarding Perceptually Salient Items. *Review of philosophy and psychology, 10*(4), 735-746. <https://doi.org/10.1007/s13164-018-0430-3>

Shapiro, K.L., et al. (1997). Personal names and the attentional blink: a visual "cocktail party" effect. *J Exp Psychol Hum Percept Perform, 23*(2), 504-514. <https://doi.org/10.1037//0096-1523.23.2.504>

Stoeber, J., & Eysenck, M.W. (2008). Perfectionism and efficiency: Accuracy, response bias, and invested time in proof-reading performance. *Journal of research in personality, 42*(6), 1673-1678. <https://doi.org/10.1016/j.jrp.2008.08.001>

Sui, J., et al. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of experimental psychology. Human perception and performance, 38*(5), 1105-1117. <https://doi.org/10.1037/a0029792>

Sui, J., & Humphreys, G.W. (2013, Nov). Self-referential processing is distinct from semantic elaboration: Evidence from long-term memory effects in a patient with amnesia and semantic impairments. *Neuropsychologia, 51*(13), 2663-2673. <https://doi.org/10.1016/j.neuropsychologia.2013.07.025>

Sui, J., & Humphreys, G.W. (2017). The self survives extinction: Self-association biases attention in patients with visual extinction. *Cortex, 95*, 248-256. <https://doi.org/10.1016/j.cortex.2017.08.006>

Sui, J., et al. (2015, Apr). The salient self: the left intraparietal sulcus responds to social as well as perceptual-salience after self-association. *Cereb Cortex, 25*(4), 1060-1068. <https://doi.org/10.1093/cercor/bht302>

Sui, J., et al. (2016). Negative mood disrupts self- and reward-biases in perceptual matching. *Q J Exp Psychol, 69*(7), 1438-1448. <https://doi.org/10.1080/17470218.2015.1122069>

Symons, C.S., & Johnson, B.T. (1997, May). The self-reference effect in memory: a meta-analysis. *Psychological Bulletin, 121*(3), 371-394. <https://doi.org/10.1037/0033-2909.121.3.371>

Team, R.D.C. (2010). R: A language and enviornment for statistical computing. In R Foundation for Statisticial Computing.

Wiecki, T.V., et al. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics, 7*, 14. <https://doi.org/10.3389/fninf.2013.00014>

Yankouskaya, A., et al. (2020). Intertwining personal and reward relevance: evidence from the drift-diffusion model. *Psychol Res, 84*(1), 32-50. <https://doi.org/10.1007/s00426-018-0979-6>

汪向东, et al. (1999). 心理卫生评定量表手册. *中国心理卫生杂志社, 13*(1), 31-35.

王力, et al. (2007). PANAS-X总维度量表在中国人群中的因素结构. *中国临床心理学杂志, 15*(6), 565-568.

王振, et al. (2011). 贝克抑郁量表第2版中文版在抑郁症患者中的信效度. *中国心理卫生杂志, 25*(6), 476-480.

1. Based on the average effect size of group-level SPE reported by Sui et al. (2012), G\*Power (f = .40, α = .05, power = 80%) revealed a minimal requirement of 16 participants. Thus, the sample size in the secondary data is sufficient to detect the self-prioritization effect at group-level. [↑](#footnote-ref-1)